The EMBO Journal vol.11 no.12 pp.4497–4506, 1992

SRY, like HMG1, recognizes sharp angles in DNA

Simona Ferrari[1], Vincent R.Harley[2], Andrea Pontiggia[3], Peter N.Goodfellow[2], Robin Lovell-Badge[4] and Marco E.Bianchi[1,5,6]

[1]Dipartimento di Genetica e di Biologia dei Microrganismi, Universita' di Milano, via Celoria 26, I-20133 Milano, Italy, [2]Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK, [3]Istituto Scientifico Ospedale San Raffaele, via Olgettina 60, I-20132 Milano, Italy, [4]National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK and [5]Dipartimento di Genetica e Microbiologia, Universita' di Pavia, via Abbiategrasso 207, I-27100 Pavia, Italy

[6]Corresponding author

Communicated by S.Ottolenghi

HMG boxes are DNA binding domains present in chromatin proteins, general transcription factors for nucleolar and mitochondrial RNA polymerases, and gene- and tissue-specific transcriptional regulators. The HMG boxes of HMG1, an abundant component of chromatin, interact specifically with four-way junctions, DNA structures that are cross-shaped and contain angles of ~60 and 120° between their arms. We show here also that the HMG box of SRY, the protein that determines the expression of male-specific genes in humans, recognizes four-way junction DNAs irrespective of their sequence. In addition, when SRY binds to linear duplex DNA containing its specific target AACAAAG, it produces a sharp bend. Therefore, the interaction between HMG boxes and DNA appears to be predominantly structure-specific. The production of the recognition of a kink in DNA can serve several distinct functions, such as the repair of DNA lesions, the folding of DNA segments with bound transcriptional factors into productive complexes or the wrapping of DNA in chromatin.

Key words: chromatin/DNA bending/HMG box/protein–DNA interactions/sex determination

Introduction

The HMG box is a recently discovered DNA binding element present in several eukaryotic proteins. It was first recognized by sequence alignments of hUBF, a transcription factor for human RNA polymerase I, with HMG1, an abundant and strongly conserved component of mammalian chromatin (Jantzen *et al.*, 1990). Additional members of the HMG family have since appeared rapidly; a few are listed in Figure 1. Sequence analysis indicates that the HMG box is a stretch of ~70 amino acids, with a net positive charge and an abundance of aromatic residues and prolines. The similarity of the primary sequences is modest: no residue is absolutely conserved and just three residues show only conservative substitutions in all known HMG boxes. However, all HMG boxes probably have a similar structure; in fact several different algorithms consistently predict two long hydrophilic α helices comprising the C-terminal half of the boxes and a possible β strand with hydrophobic character at the N-terminus (Figure 1; Bianchi *et al.*, 1992a,b).

The sequence variation between HMG boxes is paralleled by the diversity of their presumed biochemical functions. One subgroup of proteins is clearly related to HMG1 and comprises structural components of eukaryotic chromatin. HMG1-like proteins are present in all eukaryotes and in all tissues of higher organisms. Their physiological function remains elusive, however, roles have been suggested in DNA replication, nucleosome assembly and transcription (reviewed by Bustin *et al.*, 1990; Bianchi *et al.*, 1992a). Another subgroup of proteins comprises general transcription factors of RNA polymerase I and mitochondrial RNA polymerases; some of these proteins, such as hUBF (Jantzen *et al.*, 1990) contain HMG boxes as moderately repeated elements. A third subgroup has recently attracted much interest, it comprises a number of fungal proteins involved in mating type expression (Kelly *et al.*, 1988; Staben and Yanofsky, 1990; Sugimoto *et al.*, 1991), the mammalian testis determining factor SRY (Gubbay *et al.*, 1990; Sinclair *et al.*, 1990) and a set of lymphoid-specific enhancer binding factors (Travis *et al.*, 1991; van de Wetering *et al.*, 1991; Waterman *et al.*, 1991). These proteins are most probably transcriptional regulators: protein LEF-1, for example, is expressed specifically in pre-B and T lymphocytes and is involved in the transcription of the gene for the T cell receptor α chain (Travis *et al.*, 1991; Waterman *et al.*, 1991).

HMG boxes from all three subgroups have been shown to be necessary and sufficient to bind DNA: the two HMG boxes of HMG1 have been produced in *Escherichia coli* as separate polypeptides, which bind to DNA with about the same affinity and specificity as the full-length protein (Bianchi *et al.*, 1992b); *Xenopus* UBF binds to DNA via its boxes (McStay *et al.*, 1991); some mutations in the HMG box of human SRY can abolish its capacity to bind to DNA as well as its biological activity (Berta *et al.*, 1990; Harley *et al.*, 1992) and the isolated HMG box of LEF-1 retains the DNA binding properties of the whole protein (Giese *et al.*, 1991). HMG boxes are therefore authentic DNA binding domains, which can fold independently of the rest of the polypeptide.

The nature of the DNA targets recognized by HMG boxes, however, is not obvious. The transcriptional regulators (subgroup 3) produce specific footprints on DNA, spanning sequences with a recognizable consensus (Waterman and Jones, 1990; Nasrin *et al.*, 1991; Travis *et al.*, 1991). Human SRY can recognize synthetic duplex DNA fragments of the sequence AACAAAG (Harley *et al.*, 1992). Methylation interference and base substitution experiments show that LEF-1 and the related TCF-1 proteins recognize the AACAAAG motif predominantly through minor groove

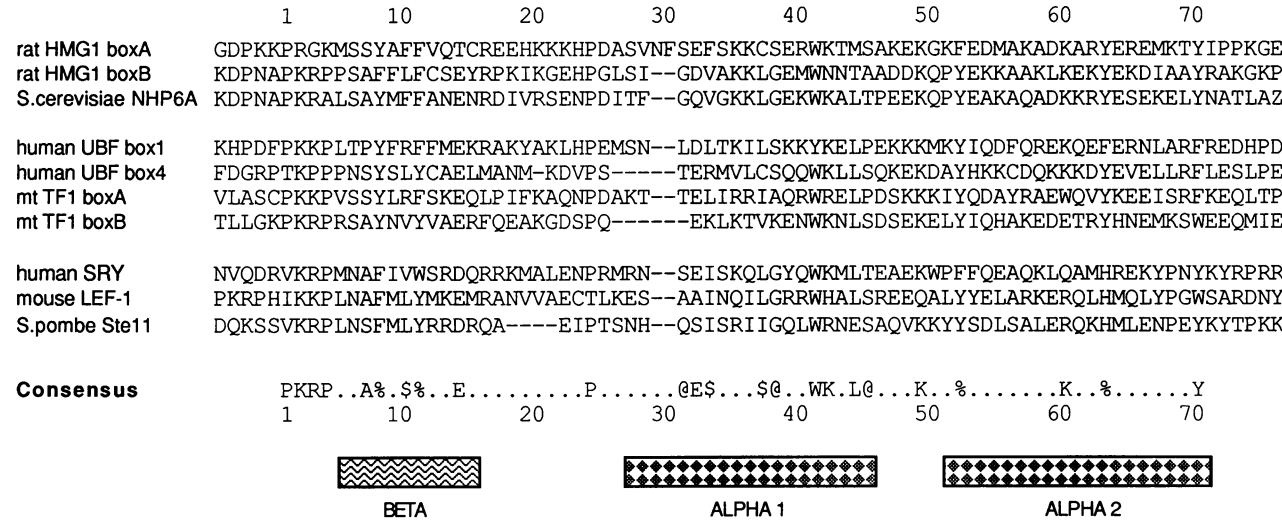

Fig. 1. Alignment of HMG boxes. Only proteins mentioned in the main text are shown; a more exhaustive compilation may be found in Bianchi *et al.* (1992a). The first group of HMG boxes are from chromatin proteins. HMG1 proteins of mammals have two HMG boxes; plant, protozoan and yeast HMG1-like proteins contain a single HMG box. The second group of HMG boxes are from general transcription factors for RNA polymerase I (human UBF, with four boxes) and mitochondrial RNA polymerases (human mtTF1, with two boxes). The third group of HMG boxes are from transcriptional regulators. Dashes indicate gaps in the alignment; Z indicates a stop codon in the gene. The consensus sequence for the HMG box motif was obtained from 21 protein sequences: one-letter symbols indicate amino acids present in 50–90% of the sequences; conservative substitutions (at least 75% of the occurrences at a particular position) are indicated as follows: @ for proline, alanine, glycine, serine and threonine; % for tryptophan, phenylalanine and tyrosine; $ for methionine, valine, leucine and isoleucine. The numbering system starts from the first conserved proline and is based on HMG1 box A. Several different computer programs predicted with a high level of confidence the two α-helices and the likely occurrence of the β-strand as shown.

contacts (van de Wetering *et al.*, 1991; Giese *et al.*, 1991, 1992; van de Wetering and Clevers, 1992). In general, all the binding sites for the HMG box transcriptional regulators are AT-rich and the same sequences are recognized by several proteins of this group. Although fairly low, the sequence specificity of each individual protein is compatible with a function in gene-specific transcription regulation.

The HMG box proteins of subgroup 2 (the nucleolar and mitochondrial transcription factors) produce specific footprints, but the protected sites do not have a recognizable consensus sequence (Bell *et al.*, 1989; Fisher *et al.*, 1989; Pikaard *et al.*, 1990a,b; Parisi and Clayton, 1991). Therefore the binding to DNA does not seem to depend entirely on sequence recognition.

The HMG box chromatin proteins (subgroup 1) are even more unusual, in that they seem to be indifferent to DNA sequence information. HMG1 protein, on the other hand, appears to have considerable affinity and specificity towards four-way DNA junctions and cruciform structures, such as those extruded from inverted repeat sequences under the effect of supercoiling (Bianchi, 1988, 1991; Bianchi *et al.*, 1989). HMG-T, the trout equivalent of mammalian HMG1, also binds to cruciform DNA (Wright and Dixon, 1988); the same is most probably true for all the members of this subgroup. The structure specificity resides in both individual HMG boxes of HMG1, although the selectivity may be somewhat relaxed in comparison to the whole protein (Bianchi *et al.*, 1992b).

HMG boxes then appear to be a class of DNA binding domains in which some members detect mainly structural cues from DNA, while others detect mainly sequence cues. This may mean that the different subclasses of HMG boxes are structurally diverse and interact with DNA in mechanistically diverse ways. The work reported here was designed to test the alternative possibility that HMG boxes recognize primarily distorted DNA and that sequence specificity may be an additional capability present in varying degrees in the three subclasses. Our results, together with the recent demonstration that LEF-1 can bend DNA almost back on itself (Giese *et al.*, 1992) and a welter of other observations discussed in a commentary by Lilley (1992), confirm that HMG boxes have a function in the recognition or generation of angles in DNA.

Results

The experimental design

Our goal is to understand the overall mechanics of interaction of HMG boxes with DNA. While detailed understanding requires the study of a large number of individual HMG boxes, the most general features of the interaction can be deduced by comparative analysis of highly diverged members of the group. HMG1 is the most typical representative of the HMG box chromatin proteins, SRY is an important factor in sex determination and biologically relevant mutations in its HMG box have previously been identified (Berta *et al.*, 1990; Harley *et al.*, 1992). We therefore chose to compare HMG box A of rat HMG1 (HMG1bA, amino acids −8 to 81 in the numeration of Figure 1) with the HMG box of human SRY, identified on the basis of its sequence similarity to other HMG boxes (Sinclair *et al.*, 1990; Figure 1). We obtained DNA fragments coding for the HMG box of human SRY by PCR of total human DNA with specific oligonucleotides (see Materials and methods). To allow the synthesis of the box in *E.coli*, valine −4 (in the numeration of Figure 1) was changed to methionine to provide a translation start site and Lys77 was changed to a stop codon. The peptide, which we called hSRYbox, was produced efficiently, was soluble and was purified to homogeneity.

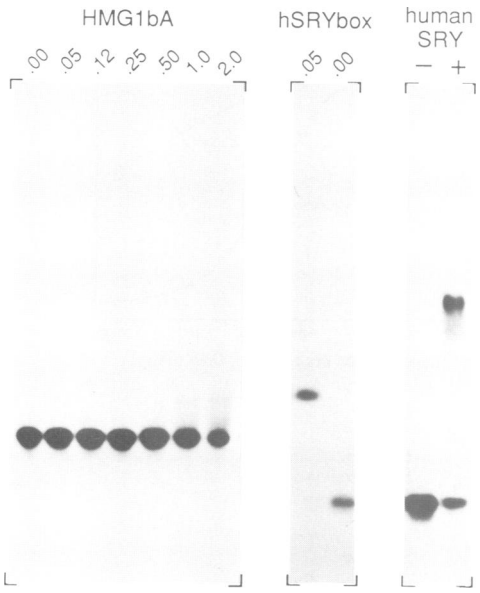

Fig. 2. HMG1bA does not recognize the AACAAAG sequence, while human SRY and hSRYbox do. Labelled probe HSS (~1 nM) was mixed in standard DNA binding buffer with purified HMG1bA, hSRYbox (concentrations of the polypeptides are indicated above the lanes as μmol/l) or unfractionated extracts from *E.coli* DH5α heat-induced cells containing the pJLA-huSRY plasmid (+ human SRY) or containing the control pJLA503 plasmid (− human SRY) (Harley *et al.*, 1992). The total protein concentration in the experiments with unfractionated cell extracts was ~500 μg/ml.

The DNA binding target: sequences versus structures

If HMG1bA and hSRYbox interact with DNA in a similar way, they should either recognize sequences on linear B-form DNA as for SRY or similar non-linear structures as for HMG1.

We first verified that HMG1bA does not recognize the HSS probe, which contains the sequence AACAAAG and was previously shown to be a preferred binding target for human SRY protein (Harley *et al.*, 1992). Both full-length human SRY and hSRYbox produce distinctive bandshifts with the HSS probe (Figure 2): this shows that the non-box regions of SRY are not essential for the binding specificity and allows us to use full-length SRY and hSRYbox interchangeably in DNA binding studies. In contrast, no specific HMG1bA−HSS probe complex is formed with comparable amounts of HMG1bA. Full-length SRY and hSRYbox, under similar conditions, do not form complexes with the control duplex HSS-allmut (Harley *et al.*, 1992), in which the AACAAAG sequence has been replaced by CCGCGGT (data not shown). We conclude that HMG1bA does not interact appreciably with the HSS probe, while SRY and hSRYbox interact with this probe in a sequence-specific manner.

The amino acid sequences of HMG1bA and hSRYbox, however, may have sufficiently diverged to route recognition towards completely different linear DNA sequences. The reverse experiment, showing that both HMG1bA and hSRYbox recognize similar structural features in DNA, was more revealing. Both HMG1bA and hSRYbox formed well defined complexes with our four-way junction probe *c* (Figure 3). The main difference between the two polypeptides is that hSRYbox can form multiple retarded bands at high protein to DNA ratios, while HMG1bA can

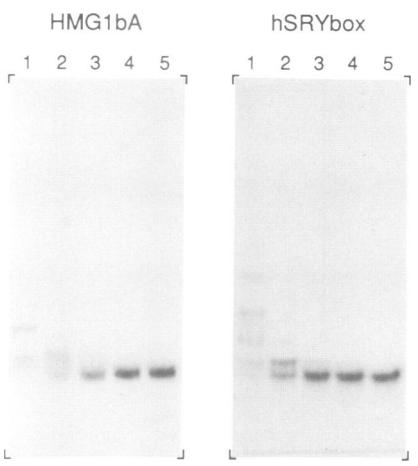

Fig. 3. Peptides HMG1bA and hSRYbox both recognize four-way junction DNA. About 0.5 ng of peptides HMG1bA and hSRYbox (lanes 1−4) or control buffer with no protein (lane 5) were mixed in standard binding buffer with various concentrations of four-way junction DNA *c*, and assayed by gel electrophoresis as described in Materials and methods. The concentration of DNA was as follows: lane 1, 1.5 nM; lane 2, 8 nM; lane 3, 40 nM; lanes 4 and 5, 200 nM,

form only two retarded bands, even at very high protein concentrations. However, the mobility of the fastest moving protein−DNA complex is strikingly similar in both cases, and in turn to the mobility of the complexes formed by full-length HMG1 (Bianchi *et al.*, 1992b). This suggests that the shape and charge of such complexes are similar, as are their calculated dissociation constants (between 10^{-9} and 10^{-8} M, data not shown). The slower moving complexes formed by hSRYbox probably contain multiple copies of polypeptide per DNA molecule, some bound to the high affinity sites at the base of the junction and some to low affinity sites on the arms of the junction.

The binding affinities of hSRYbox for linear and junction DNA were examined in more detail by means of competition assays. Figure 4A shows that hSRYbox is indifferent to the specific sequences present in four-way junctions. Both junction *c* and an additional four-way junction of unrelated sequence (junction 1, Duckett *et al.*, 1988) compete with the labelled HSS probe better than cold HSS DNA fragments. This observation was confirmed by using yet another four-way junction, junction *f* (data not shown). None of the three junctions contain sequences overtly related to the AACAAAG proposed binding site for SRY.

We showed previously that the selectivity of HMG1bA towards four-way junctions is formidable: linear DNAs in 1000-fold excess do not compete detectably for the binding (Bianchi *et al.*, 1992b). This is consistent with a very low affinity of HMG1bA for linear DNA of whatever sequence. We found, however, that linear DNAs *a* and *b*, containing the same sequences as the four-way junction *c*, compete appreciably with the labelled junction probe for binding to hSRYbox, even if less efficiently than cold junctions (data not shown). These results are consistent with a fairly high non-specific binding affinity of hSRYbox for linear DNA. Our interpretation is supported by other observations. Although LEF-1 forms a very stable complex with a specific sequence present upstream of several genes expressed in T-cells, the affinity for aspecific competitor linear DNA is only 50-fold lower than that for the specific target (Giese

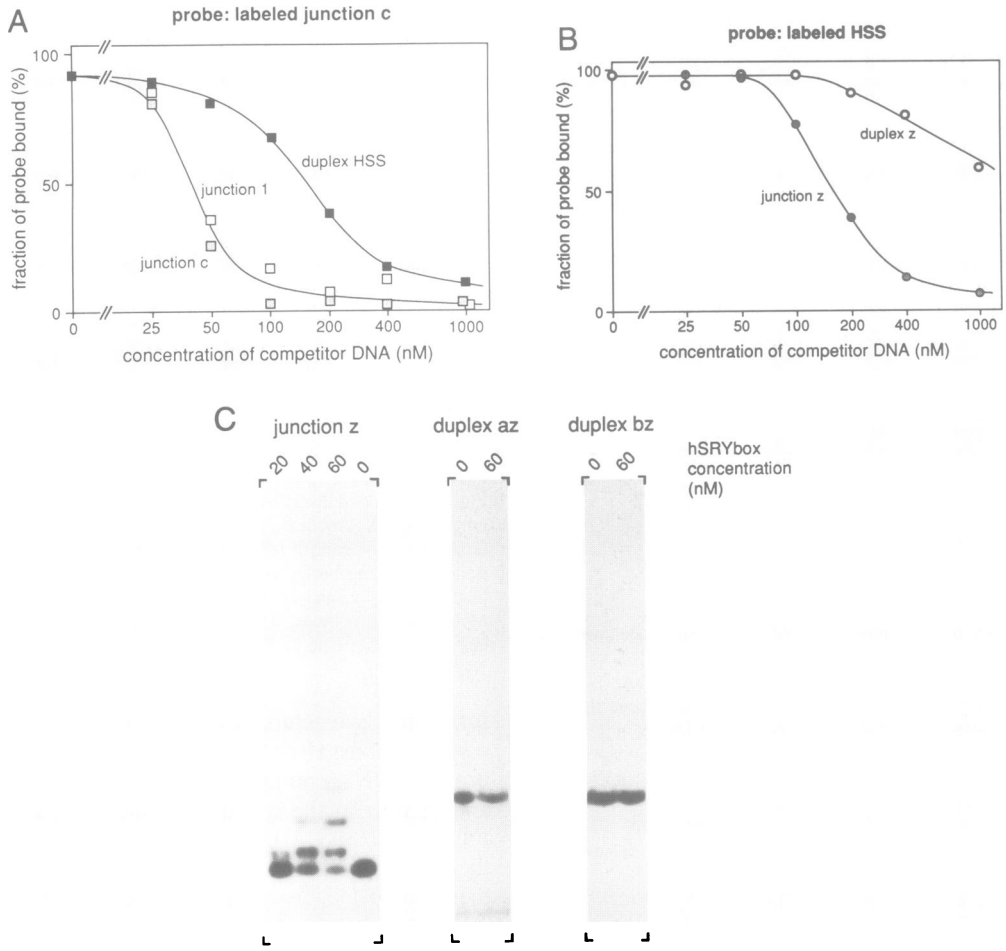

Fig. 4. hSRYbox recognizes four-way junctions structure-specifically and not sequence-specifically. **A**. Two different four-way junction DNAs are recognized by peptide hSRYbox with similar efficiency. About 5 ng of peptide hSRYbox were mixed in a total volume of 10 μl of standard binding buffer containing 1 nM of labelled four-way junction *c* and the indicated amounts of unlabelled four-way junction *c* (grey squares), junction 1 (Duckett *et al.*, 1988) (open squares) or HSS duplex DNA (black squares). Electrophoresis, autoradiography and quantification of the autoradiographic signals were done as described in Materials and methods. **B**. Peptide hSRYbox recognizes preferentially the specific structure of four-way junctions. To discriminate between structure-specific and sequence-specific binding, four-way junction *z* was constructed in such a way to contain only sequences that interact weakly with SRY. The two control duplex DNAs (*az* and *bz*, indicated collectively in the figure as duplex *z*) contain all the sequence information present in junction *z*, but are linear (see Materials and methods). About 5 ng of peptide hSRYbox were mixed in a total volume of 10 μl of standard binding buffer containing 1 mM of labelled HSS duplex DNA and the indicated amounts of unlabelled four-way junction *z* (grey circles) or of the two duplexes (open circles). **C**. Peptide hSRYbox does not recognize linear control DNAs *az* and *bz* at the concentrations optimal for complex formation with junction *z*. Labelled junction *z*, duplex *az* and *bz* (0.2 nM) were mixed in 10 μl of standard binding buffer with the indicated amounts of hSRYbox peptide. Electrophoresis and autoradiography were done as described in Materials and methods.

et al., 1991). Also, single mutations in the AACAAAG binding site are often well tolerated by SRY protein (Harley *et al.*, 1992). To obtain convincing proof that hSRYbox recognizes the peculiar shape of four-way junctions and not sequence-specific binding sites adventitiously present in the junction, we synthesized junction *z*, which contains sequences deliberately chosen as poor binding sites for SRY. In this case, the linear control duplexes compete poorly with the four-way junction (Figure 4B). In addition, the hSRYbox−junction complex is formed at peptide concentrations that are completely ineffectual for the formation of hSRYbox−linear DNA complexes (Figure 4C); the latter are formed at low level (< 10% of input DNA) at ~300 nM hSRYbox concentration, again indicating that linear DNAs can be bound to some extent by this HMG box (not shown). We conclude that the 'structure versus sequence' selectivity of hSRYbox varies between a factor of four and 100, depending on the actual sequences present in the DNA, and is hence much lower than the selectivity of either HMG1bA or full-length HMG1. However, the affinity for junction DNA is somewhat higher than the affinity for the specific AACAAAG binding site in linear DNA.

SRY protein bends linear DNA

The results described in the previous sections establish that the HMG box of human SRY protein can bind efficiently and with moderate selectivity to two structurally distinct targets: strongly distorted DNA and linear DNA containing specific sequences. Peptide hSRYbox is sufficiently small to rule out the possibility that it may contain two different binding sites. Simple thermodynamic considerations suggest that if the two preferred ligands of SRY interact with its HMG box in similar ways, they should either have a similar ground-state conformation (for a lock and key mode of interaction with the peptide) or similar potential conformations that can be reached by overcoming energetic barriers not

much larger than the overall free energy of protein–ligand association (an induced-fit mode of interaction).

To test the induced-fit hypothesis, we performed a circular permutation assay to detect possible large distortions induced on linear DNA by the interaction with the HMG box of human SRY. DNA fragments with a distortion in the middle of the molecule have a different shape, and hence different electrophoretic mobility, compared with DNA fragments of identical length and composition with a distortion near one end (Wu and Crothers, 1984). Although the relation between electrophoretic mobility and conformation is complex (Levene and Zimm, 1989), measuring the rate of migration in polyacrylamide gels of complexes of protein with DNA fragments of circularly permuted sequence allows one to map the locus of protein–DNA interaction and to estimate the amount of distortion introduced in DNA (Liu-Johnson *et al.*, 1986; Thompson and Landy, 1988; see also Materials and methods).

To generate the probes for the permutation assay, a short sequence from the enhancer of the CD3ϵ gene, containing the AACAAAG sequence and previously shown to be a good binding site for human SRY protein (Harley *et al.*, 1992), was cloned between directly repeated sequences in plasmid pBend2 (Kim *et al.*, 1989). Cleavage with several restriction endonucleases yielded a set of fragments of identical length and circularly permuted sequence (Figure 5A). We incubated these fragments with unfractionated *E.coli* extracts containing full-length human SRY protein or control extracts and analysed the electrophoretic mobility of the resulting complexes (Figure 5B). The formation of complexes was dependent on the presence of SRY protein. Complexes with the CD3ϵ binding site in the middle (fragment D) migrated significantly more slowly than complexes with the site near the ends (fragments A and G). No statistically significant difference in the mobility of the free DNA probes was observed, indicating that the CD3ϵ site does not distort DNA on its own. By analysing the data in terms of a simple geometric model (Figure 5C; see also Materials and methods), we localized the site of flexure of the DNA to the centre of the AACAAAG site, and estimated a large deviation of the axis of DNA from linearity (~83°).

While this work was in progress, Grosschedl and coworkers reported a similar experiment performed with a fusion protein containing the HMG box of mouse SRY and a set of circularly permuted DNA fragments containing the sequence AACAATG, for which mouse SRY has a much higher affinity than for the AACAAAG sequence (Giese *et al.*, 1992). Their estimated deflection of 85° is extremely close to our own estimate. Although the value obtained for the bend angle depends somewhat on the algorithm used, and may not be absolutely precise, it is clear that mouse and human SRYs bend their binding sites in very similar ways.

Increased flexibility of the DNA without a specific orientation can also cause sizeable differences of electrophoretic mobility in the circular permutation assay (Gartenberg and Crothers, 1988; Kerppola and Curran, 1991a,b). However, we believe that the results obtained with human and mouse SRY are best interpreted as a sharp bend or kink introduced in DNA, both because various models for the interpretation of the electrophoretic mobility data give consistent results, and because the size of the distortion caused by SRY proteins is quite large. To explain it by increased flexibility, one should assume nearly total

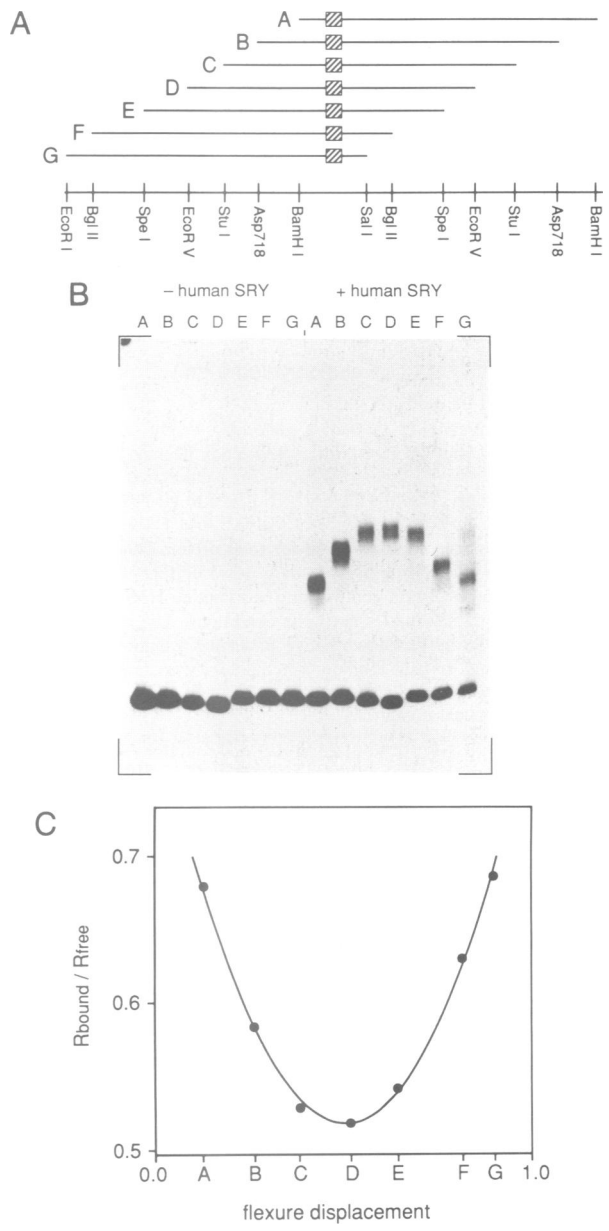

Fig. 5. Circular permutation analysis of DNA flexure induced by binding of human SRY to the AACAAAG sequence. **A.** Plasmid pB2CD3ϵ, containing the CD3ϵ site (hatched box) flanked by tandemly duplicated DNA sequences, was cleaved at the restriction sites indicated in the map. The DNA fragments obtained in this way (designated A–G) all contain circular permutations of the same sequence of 141 bp. **B.** Electrophoretic mobility of the circularly permuted DNA fragments complexed to the human SRY protein. DNA fragments A–G (8 fmol) were mixed with sonicated salmon sperm DNA (50 ng) in 9 μl of standard DNA binding buffer (see Materials and methods). To the various mixtures 1 μl of unfractionated *E.coli* extract (~5 μg of total protein) containing human SRY or the same volume of control extract was added. Electrophoresis and autoradiography were performed as indicated in Materials and methods. **C.** Mapping of the locus of flexure and analysis of the bending parameters. The mobilities of the protein–DNA complexes (R_{bound}) were normalized to the mobility of the corresponding free DNA (R_{free}). The distance of the centre of the CD3ϵ site from the 5′ end of the probe was divided by the total length of the probe (flexure displacement). The plotted points were interpolated with a quadratic function as described in Materials and methods. The fitting second-order equation was $y = 1.117x^2 - 1.131x + 0.804$ ($R^2 = 0.997$). The first- and second-order parameters of the equation are in close agreement and yield an estimate of deviation from linearity of ~83°. The locus of flexure was localized to the centre of the AACAAAG sequence, ±2 bp.

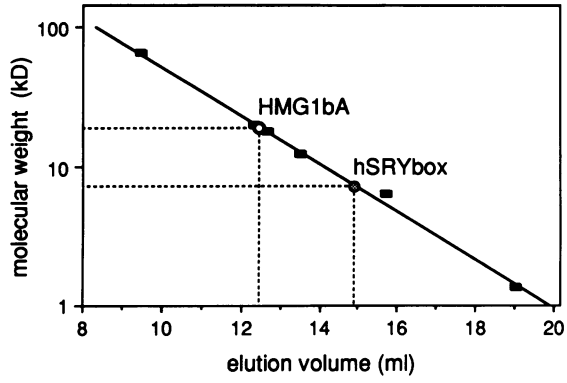

Fig. 6. Determination of the native molecular weight of hSRYbox by gel filtration. About 5 μg of purified hSRYbox peptide were applied to an FPLC Superdex 75 column (Pharmacia) equilibrated with buffer D-500 (containing 20 mM HEPES, pH 7.9, 500 mM NaCl, 0.2 mM EDTA, 0.05% NP40, 10% glycerol and 0.5 mM DTT). Fractions of 300 μl were collected and analysed for total protein content (by SDS–PAGE) and DNA binding activity towards four-way junction *c* and the HSS probe. As a control, purified peptide HMG1bA (~5 μg) was chromatographed under the same conditions. The elution volumes of marker molecules of known native molecular weight (bovine serum albumin, soybean trypsin inhibitor, myoglobin, cytochrome *c*, aprotinin and vitamin B12) were determined by measuring OD_{280} nm in separate runs under identical conditions.

rotational freedom at one of the base pair steps within the site of flexure.

Peptide hSRYbox is a monomer in solution

We determined previously that both HMG boxes of HMG1, HMG1bA and HMG1bB, behave as dimers in solution (Bianchi *et al.*, 1992b). It was therefore of interest to determine the subunit composition of native hSRYbox.

An aliquot of a purified preparation of hSRYbox was applied to a gel filtration FPLC column; the fractions of the eluate were assayed for protein content and DNA binding activity. In comparison with molecular weight standards, hSRYbox emerged as a peak centred around an M_r of 7.3 ± 0.4 kDa (Figure 6). As a control, the activity of HMG1bA emerged at 19 ± 0.2 kDa. The molecular weights calculated from the sequence of the genes expressed in *E.coli* is 9.8 kDa for hSRYbox and 10.6 kDa for HMG1bA. It therefore appears that hSRYbox is a monomer in solution, while HMG1bA is a dimer. Of course, the difference in elution profiles might also be attributed to a radically different shape of the two HMG boxes, but this seems unlikely. Giese *et al.* (1991) concluded from electrophoretic mobility shift assays that LEF-1 binds DNA as a monomer; similar experiments indicated a monomeric structure for full-length SRY (V.R.Harley, unpublished results). We have at present no simple way to reconcile the different subunit composition of HMG boxes with HMG1 and SRY, however, it is interesting to note that HMG boxes are duplicated in tandem within HMG1, whereas they are present in single copy within all the HMG box transcriptional regulators, thus reproducing the stoichiometry we found for free boxes.

Discussion

The binding of HMG boxes to DNA

HMG boxes are DNA binding domains originally identified on the basis of primary sequence similarity in HMG1 protein, general transcription factors and gene- and tissue-specific transcriptional regulators. HMG1 and HMG1-like proteins are very abundant components of eukaryotic chromatin and their function is still rather obscure (Bustin *et al.*, 1990). We found recently that HMG1 has a peculiar specificity in DNA binding: it recognizes four-way junctions, DNA structures that can be generated by recombination events and by intrastrand base pairing of inverted repeat sequences (Bianchi *et al.*, 1989). The binding specificity of HMG1 is imparted onto it by its HMG boxes (Bianchi *et al.*, 1992b). Although four-way junctions have been shown to exist in both prokaryotic and eukaryotic cells (Bell and Byers, 1979; Panayotatos and Fontaine, 1987; Horwitz and Loeb, 1988), their occurrence is certainly rather rare and it is difficult to imagine why abundant, non-enzymatic nuclear proteins should have evolved to recognize them. An additional puzzle was that another group of HMG box proteins, including the lymphoid enhancer-binding protein LEF-1 (Travis *et al.*, 1991) and the testis determination factor SRY (Sinclair *et al.*, 1990), appeared to be classical sequence-specific binding proteins, albeit with lower sequence specificity than most transcriptional regulators (Giese *et al.*, 1991; Nasrin *et al.*, 1991; Harley *et al.*, 1992). Yet another group of HMG box proteins, including the RNA polymerase I transcription factor UBF and the mitochondrial polymerase transcription factor mtTF1, bind to DNA in specific positions 5′ to the transcription start site, but do not seem to recognize simply a specific succession of bases: no consensus sequence can be discerned (Bell *et al.*, 1989; Fisher *et al.*, 1989; Pikaard *et al.*, 1990a,b). Our experiments attempt at reconciling this set of observations into a coherent model.

We have first shown that binding to four-way DNA junctions is not a property unique to the HMG boxes of HMG1, but on the contrary is enjoyed also by the HMG box of SRY. The stabilities of the complexes formed with four-way junctions by HMG1bA and hSRYbox are comparable and do not appear to depend on the specific sequences that make up the junctions: three different junctions were bound with approximately the same efficiency. We have then established a connection between the superficially contradictory abilities of the HMG box of SRY: to be able to recognize a fairly specific succession of bases (Harley *et al.*, 1992) and yet be able to recognize four-way junctions irrespective of sequences. The reciprocal competitive inhibition of binding of the HSS probe and junction *c* suggested that these two different DNA molecules had a potentially similar structure, at least when complexed to SRY. The circular permutation assays carried out by Giese *et al.* (1982) on mouse SRY and by us on human SRY confirm this prediction: the bound DNA fragments contain an angle of nearly 90°. Four-way junctions have a folded structure resembling an X: the four arms lie almost in a plane, forming two angles of ~120° and two angles of ~60° (Murchie *et al.*, 1989; Lilley, 1990; Bhattacharyya *et al.*, 1991). If we consider any two arms of the junction, they resemble a linear DNA with an angle, such as the one contained in DNA bound by SRY proteins. At present, it is difficult to say whether HMG boxes recognize preferentially an obtuse or an acute angle in the junction and whether they modify its amplitude on binding. However, the HMG box of LEF-1 protein produces an included angle of ~50° (Giese *et al.*, 1992). It is quite probable that it will recognize preferentially one acute angle in the junction. By analogy,

we consider a 60° angle in the junction as the most likely binding site for all HMG boxes.

On a more general level, the demonstration that the HMG box of a sequence-specific DNA binding protein recognizes kinked DNA irrespective of its sequence supports the notion that the interaction with distorted DNA is the general property of HMG boxes, central to their biological function (Bianchi *et al.*, 1992b). Lilley (1992) has suggested that HMG box domains have been recruited in functionally diverse proteins to perform various manipulations of DNA structure. Such contortions can be required for DNA transcription, repair and packaging; the function of each individual protein will depend on the other domains that cooperate with the HMG box. As examples, LEF-1 protein may be involved in the spatial organization of the nucleoprotein complex required for DNA transcription (Giese *et al.*, 1992); another recently cloned HMG box protein is probably involved in the repair of DNA distorted by the presence of cisplatin adducts (Bruhn *et al.*, 1992). HMG1 may serve a generalized 'architectural' role in DNA bending/looping/folding/wrapping, of which we have witnessed so far only a few manifestations.

Mechanistic implications

Although all HMG boxes recognize similar structural features in DNA, whatever the specific function of the protein that contains them, the HMG boxes of HMG1 and SRY are not interchangeable. A first difference is that HMG1bA forms dimers in solution while hSRYbox is monomeric. More significantly, HMG1 recognizes only distorted DNA and does not bind to linear DNA; the DNA recognized by SRY and LEF-1 is initially linear and becomes bent only when complexed to the protein.

The interaction of HMG boxes of the SRY/LEF-1 type with linear DNA probably occurs in two steps: the initial, sequence-specific recognition and the formation of the complex with bent DNA, which is largely sequence-insensitive. The favourable contacts between DNA and protein that promote the recognition step may not be exactly the same as those that promote the stability of the final complex.

The interaction of LEF-1 and SRY with linear DNA has been shown to proceed mainly through the minor groove, as revealed by methylation interference experiments and substitution of hydrogen bond donors and acceptors in the major and the minor grooves (Giese *et al.*, 1991, 1992; van de Wetering and Clevers, 1992). Such a mode of interaction is unusual for sequence-specific DNA binding proteins and is probably responsible for the fairly modest specificity of LEF-1 and SRY for their binding sites (Giese *et al.*, 1991). The minor groove provides small opportunity for base-specific contacts, since hydrogen bonding cannot distinguish T from C (Starr and Hawley, 1991) or AT pairs from TA pairs (Seeman *et al.*, 1976). However, hydrogen bonding in the minor groove appears well suited for structure-directed recognition, since the phosphates will be spaced at favourable distances for selective interactions. Conversely, it appears quite probable that proteins which contact the DNA through the minor groove will display a pronounced structure specificity. Two such proteins, *E.coli* Integration Host Factor (IHF) (Yang and Nash, 1989) and the TATA-binding protein TFIIDτ (Lee *et al.*, 1991; Starr and Hawley, 1991) in fact bend DNA, the former by a dramatic 140° (Thompson and

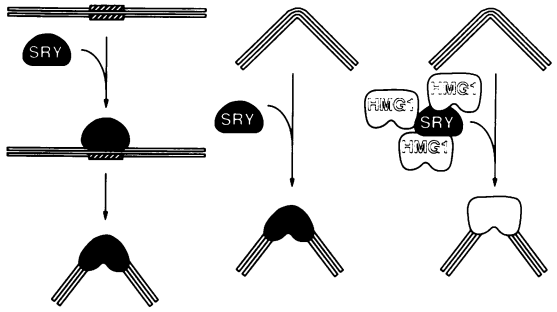

Fig. 7. A model for the interaction of HMG box proteins with DNA. We argue that human SRY protein can bind to linear DNA (left cartoon); binding to specific sequences (hatched) produces a stronger interaction either because of specific hydrogen bonds made to the bases within the specific site, or because the specific site offers lower resistance to structural deformation. The interaction of SRY with the specific site brings about a rearrangement in the structure of both SRY and the DNA, which becomes bent. When DNA which is already distorted (because of its composition or because of DNA damage, or due to interaction with other proteins as in looping) becomes available to SRY, SRY binds with no sequence discrimination (middle cartoon). The binding of such proteins sensitive to DNA conformation is predicted to be very sensitive to the chromatin environment. HMG1-like proteins recognize with high affinity distorted DNA; since their concentration is orders of magnitude higher than the concentration of SRY and other distortion-seeking proteins, they will occupy the distorted sites with higher probability (right cartoon). In so doing they will restrict the binding of proteins like SRY to the sequence-specific sites.

Landy, 1988; Horikoshi *et al.*, 1992). In addition, we have found that protein HU, a 'histone-like' protein ubiquitous in prokaryotes (Drlica and Rouvière-Yaniv, 1987) and very similar to IHF, can bind four-way junction DNA structure-specifically and with no sequence preferences (A.Pontiggia and M.E.Bianchi, unpublished results). It will be interesting to determine whether IHF and TFIIDτ can do so as well.

The deflection of a linear double helix into a sharp angle requires an extensive rearrangement of the DNA geometry. A small number of bases will have to be repositioned, leading to a very significant local alteration in the width of the grooves and at least the unstacking of one base-pair step or partial unstacking of several steps. The energetic cost of such a conformational modification of DNA must be high and must be paid out of the free energy liberated by more extensive and/or intimate contacts in the final complex, compared with the initial one. The protein–DNA contact surface cannot be very large in the case of hSRYbox, due to the small size of the polypeptide involved; we suspect that even a very snug fit in the complex will not be sufficient to provide the free energy required for DNA kinking. We suggest that the structure of the HMG box domain will also have to be rearranged in the transition between the initial and final complexes (depicted schematically in Figure 7), so that a better packing within the domain will also contribute to the stability of the protein–DNA complex. Such an induced-fit type of interaction has been proposed for the binding of Jun and Fos proteins with the AP-1 site (Kerppola and Curran, 1991a,b). The conformational switch in this case involves primarily the basic region adjacent to the leucine zippers, which is poorly structured in solution and is stabilized into an α-helix with packing surfaces when DNA is bound. Peptide hSRYbox contains extensive α-helices and two tryptophans, and the transition we propose may

be amenable to investigation by circular dichroism or fluorescence studies.

The interaction of hSRYbox with four-way junctions must closely mimic the interaction in the final complex just described. In particular, the lack of obvious sequence preferences in the binding of hSRYbox to four-way junctions suggests that specific hydrogen bonding between amino acids and bases cannot contribute decisively to this interaction, which should then depend mainly on precise van der Waals contacts and non-base-specific hydrogen bonds, most probably to the DNA backbone. In fact, four-way junctions already contain the sharp angle which seems to be an intrinsic feature of DNAs bound to HMG boxes.

In this context, the inability of HMG1bA to bind efficiently to linear DNA suggests that the HMG1bA polypeptide is less flexible and unable to accommodate the full swing between linear and sharply bent DNA. In this case the interaction with DNA would approximate a lock and key model. The observation that HMG1 can bind to DNA containing a kink of ~40° (Pil and Lippard, 1992), however, suggests that the variety of distortions accommodated by HMG1 is sufficiently wide. The HMG boxes of UBF and related proteins, which recognize their binding sites in linear DNA, but apparently not through base-specific contacts, may represent yet another variation on this basic theme.

We should also underline that our results suggest one possible function for HMG1-like proteins, though probably not the only one. Due to their much higher concentration, HMG1-like proteins should be able to saturate DNA sites that for a variety of reasons are transiently bent, acting as decoys for distortion-seeking proteins and causing their misappropriate binding (Figure 7). Such a role would be entirely consistent with the modest but reproducible stimulatory effect of HMG1 and 2 on transcription by RNA polymerases II and III (Singh and Dixon, 1990; Tremethick and Molloy, 1986, 1988) and with their role in facilitating the binding of the specific transcription factor MLTF (Watt and Molloy, 1988).

Materials and methods

Construction of synthetic junction and linear DNAs

Oligonucleotides were synthesized by the phosphotriester method and purified by HPLC. Duplex and four-way junction DNA molecules were obtained by annealing the appropriate oligonucleotides and were purified by 6.5% PAGE.

Four-way junction *c* is composed of four strands of 30, 35, 40 and 46 nucleotides. Junction *f* is composed of four strands of 46, 47, 50 and 50 nucleotides, without any sequence similarity to those composing junction *c*. As controls for structure-specific binding, two linear duplex DNAs were used, called *a* and *b*. Molecule *a* is composed of the 35 nucleotide strand of junction *c* annealed to its antiparallel complement; molecule *b* is composed of the 40 nucleotide strand of junction *c* annealed to its antiparallel complement. Therefore molecules *a* and *b* together have the same sequences present on junction *c*, the same number of double-stranded ends, the same approximate mass, but a different three-dimensional structure. Details of the construction of these molecules are given in Bianchi *et al.* (1989).

Four-way junction *z* is composed of the four 30mers ol1 AGCGCTCTCACACGGGCCTCCGCCCAGCTG, ol3 CAGCTGGGCGGAGGGCGGACGTTAACCCC, allmut.dir GGGGTTAACGTCCGCGGTAATCTGGTAGA, ol4 TCTACCAGATTACCCCCGTGTGAGAGCGT. Control duplex *az* was constructed by annealing ol1 with its complement ol2 (CAGCTGGGCGGAGGCCCGTGTGAGAGCGCT); the sequences contained in the resulting duplex were shown to be poor binding sites for TCF-1 (van de Wetering *et al.*, 1991). Control duplex *bz* was constructed by annealing ol allmut.dir with its complement ol allmut.rev; the resulting duplex is identical to the allmut probe, which was shown previously to be a poor binding site for SRY (Harley *et al.*, 1992).

The HSS probe is identical to the HuSRY duplex DNA described by Harley *et al.* (1992) and consists of the annealed oligonucleotides GGGGTTAACGTAACAAAGAATCTGGTAGA and TCTACCAGATTCTTTGTTACGTTAAC, labeled with the Klenow fragment of DNA polymerase and [α-^{32}P]dCTP.

Construction of plasmid pT7-hSRYbox

Oligonucleotides hSRYboxdir (CCACATATGCAGGATAGAGTGAAGCGA) and hSRYboxrev (CGAAGCTTAACGACGAGGTCGATACTT) were synthesized by the phosphotriester method and were used for PCR without purification. PCR mixtures (50 μl) contained 50 pmol each of oligonucleotides hSRYboxdir and hSRYboxrev, 0.2 mM dNTPs, 400 ng purified human total genomic DNA, 1 unit *Taq* DNA polymerase and 5 μl *Taq* polymerase 10 × buffer (Perkin–Elmer Cetus). Thirty cycles of denaturation (30 s at 94°C), annealing (60 s at 50°C) and polymerization (90 s at 72°C) were performed on a Perkin–Elmer Cetus instrument. Reactions 2 and 3, using DNA from two different male students, yielded ~1 μg of a single PCR product ~0.25 kb long. Reaction 1, using DNA from a female student, yielded no visible product. The product of reaction 3 was cleaved with restriction endonucleases *Nde*I and *Hin*dIII, gel purified and cloned between the *Nde*I and *Hin*dIII sites of plasmid pT7-7 (Tabor and Richardson, 1985). The resulting plasmid, pT7-hSRYbox, was checked by sequencing with T7 DNA polymerase and was then introduced in strain BL21(DE3) (Studier *et al.*, 1991).

Construction of plasmid pB2CD3ϵ and probe preparation

Plasmid pB2CD3ϵ was prepared by insertion of the annealed synthetic oligonucleotides CTAGAGAGCGCTTTGTTCTCAG and TCGACTGAGAACAAAGCGCTCT between the *Xba*I and *Sal*I restriction sites in plasmid pBend2 (Kim *et al.*, 1989). DNA probes for electrophoretic mobility assays were prepared by restriction enzyme cleavage of pB2CD3ϵ and purified by agarose gel electrophoresis. The probes were then labelled by filling in protruding ends with the Klenow fragment of DNA polymerase, [α-^{32}P]dATP, dCTP, dGTP and TTP. When no protruding ends were generated, the probes were dephosphorylated with calf intestinal phosphatase and labelled with T4 polynucleotide kinase and [γ-^{32}P]ATP. The probes were adjusted to the same concentration and specific activity.

Preparation of cell extracts and purification of peptides HMG1bA and hSRYbox

Full-length human SRY protein was synthesized in *E.coli*. Cells of the strain DH5α, either bearing plasmid pLAJ503-huSRY or the control plasmid pLA503, were grown, heat-shocked and lysed as described by Harley *et al.* (1992).

Peptide HMG1bA was prepared as described by Bianchi *et al.* (1992b). Peptide hSRYbox was synthesized in BL21(DE3) cells harbouring plasmid pT7-hSRYbox. Cells were grown, induced, harvested and sonicated as described for the preparation peptide HMG1bA. Nucleic acids were removed by batch absorption to DEAE–cellulose at 0.45 M NaCl. The extract was mixed on ice with solid ammonium sulfate (to 2 M) and centrifuged at 10 000 r.p.m. in an SS34 rotor for 20 min at 0°C. The supernatant was mixed with additional ammonium sulfate (to 3.2 M) and centrifuged as before. The pellet was resuspended in a buffer containing 20 mM HEPES, pH 7.9, and 0.5 mM DTT, and subjected to FPLC on a MonoS column (Pharmacia). Peptide hSRYbox eluted at ~0.55 M NaCl. The peak fractions were pooled and dialysed; the preparation was completely homogeneous.

Assay for DNA binding

DNA binding buffer (10 μl of final volume) contained 8% Ficoll, 200 mM NaCl, 10 mM HEPES, pH 7.9, 5 mM KCl, 1 mM EDTA, 1 mM spermidine and 0.5 mM DTT. To these components we added in various combinations the linear DNA probes, the four-way junction probes, sonicated salmon sperm DNA and purified polypeptides or unfractionated cell extracts (final concentrations are indicated in the legends to the figures). The usual order of addition was fixed components, polypeptides and then DNAs (labelled probe plus cold competitors). In separate experiments (not shown) the order of addition was varied in order to verify that the equilibrium between free and bound ligands had been reached. After incubation for 10 min on ice, samples (5 μl) were applied to vertical 6.5% polyacrylamide gels in 0.5 × TBE and electrophoresed at 11 V/cm for 3–4 h at room temperature. The gel was then fixed in 10% acetic acid, dried and autoradiographed with Kodak XAR-5 films at −80°C for 24–72 h with intensifying screens.

Calculation of DNA bend parameters

For circular permutation analysis, the mobilities of protein–DNA complexes were normalized to the mobility of free DNA (R_{bound}/R_{free}, vertical axis

of the graph in Figure 5C). The distance between the 5′ end of the probe and the apparent centre of flexure were normalized to the total length of the probe (flexure displacement, horizontal axis of the graph in Figure 5C). The points in the graph were interpolated with a second-order equation (a parabola) by means of least squares algorithm (Cricket Graph application on a Macintosh computer).

To analyse the electrophoretic mobilities of the protein–DNA complexes, we adopted the Lumpkin–Zimm reptation model, in which the DNA chains migrate in wormlike fashion among the gel fibres. The DNA chains are confined to a tube, composed of a sequence of segments that connect the consecutive points of contact between the DNA and the gel fibres. The mobility of the chain, R, is proportional to the centre of mass velocity, v_{cm}, of the chain in the direction of the electric field such that:

$$R = <v_{cm}>/E = <h_x^2/L^2> Q/\zeta$$

(equation 1, Levene and Zimm, 1989)

where the field of strength E is along the x axis, Q is the total charge of the DNA, ζ is the friction constant for motion along the tube, h_x is the component in the x direction of the tube's end to end vector, L is the contour length of the tube and the angle brackets denote an average over an ensemble of conformations. In this model, $<h_x^2/L^2>$ can be smaller than unity for two reasons: the introduction of a fixed, oriented bend in the DNA molecule or the presence of an ensemble of non-fixed, non-oriented bends, such as those allowed by a loose hinge. Levene and Zimm (1989) have computed with Monte Carlo simulations R and $<h_x^2/L^2>$ for straight and bent chains and have compared them with the experimental electrophoretic mobilities of bent DNA molecules. They found that the results are closely approximated by the model if one introduces an additional, independently adjustable elastic force constant. B_{eff}, which accounts for the relative deformability of gel and DNA. In our analysis, we have drastically simplified this mathematical complexity by adopting the following assumptions: (i) for circularly permuted DNA chains of fixed length, Q and ζ are constant; (ii) the flexure introduced in the DNA chain by the binding of the protein is an angle, θ, with a well identified vertex and a fixed amplitude (as opposed to a continuous bend and a dynamically averaged ensemble $<\theta>$ of amplitudes) and in addition (iii) due to the limited length of our probes and the ionic strength of the gel system, DNA chains are essentially rigid, so that $<h_x^2/L^2>$ can be approximated by h_x^2/L^2; *(iv)* B_{eff} does not vary substantially with the displacement of the angle along the circularly permuted DNA molecules. Condition (ii) is probably the most critical, since situations have been found in which the protein-induced flexure resembles more closely a loose hinge than a rigid angle (Gartenberg and Crothers, 1988; Kerrpola and Curran, 1991a). Conditions (ii) and (iv) can be partially checked by running the complexes in gels of different polyacrylamide concentration; the calculated angle amplitudes should not vary significantly. Under the conditions specified, equation (1) reduces to:

$$R_{bound}/R_{free} = k_{bound} (h_{xbound}^2/L^2)/k_{free} (h_{xfree}^2/L^2) = Kh_{xbound}^2/L^2 \quad \text{(equation 2)}$$

where k (and hence K) are constants and $h_{xfree} = L$ for a straight rod. How does h_{xbound} depend on D, the distance of the vertex of the angle θ from the 5′ end of the DNA molecule? In a triangle, the length of the three sides a, b and c and the angle γ subtended by a and b are related by the formula

$$c^2 = a^2 + b^2 - 2abcos\gamma \quad \text{(equation 3)}$$

In our model

$$h_{xbound}^2 = D^2 + (L - D)^2 - 2D(L - D)cos\theta \quad \text{(equation 4)}$$

Substituting equation (4) in equation (2) yields

$$R_{bound}/R_{free} = K[D^2 + (L - D)^2 - 2D(L - D)cos\theta]/L^2 = 2K(1 + cos\theta)(D/L)^2 - 2K(1 + cos\theta)(D/L) + K \quad \text{(equation 5)}$$

Thus in our model R_{bound}/R_{free} is a quadratic function of D/L and the experimental values for R_{bound}/R_{free} can be interpolated by a parabola, whose minimum identifies the locus of flexure. In addition, the amplitude of θ can be readily derived from the parameters for the second-order and first-order terms of the equation, both equal to $2K(1 + cos\theta)$; K is the zero-order parameter of the same equation. In fact, the comparison of the two estimates of the angle θ derived from the first- and second-order parameters is a good test of the model.

A similar geometrical treatment of the problem was used by Thompson and Landy (1988). They derived the formula $\mu_M/\mu_E = cos\alpha/2$, which relates the angle of deviation from linearity of the DNA, α, to the relative mobilities of complexes with a flexure exactly in the middle, μ_M, or at the end of the molecule, μ_E. Empirically, the formula $\mu_M/\mu_E = cos\alpha/2$ was found to give a good fit to observed values for α angles between 0° and 140° (θ angles between 180° and 40°). We find that our solution gives results numerically similar to that of Thompson and Landy (1988) and is more accurate and robust, since it considers many data points rather than two (μ_M and μ_E).

Acknowledgements

We thank Luca Falciola and Raffaele De Francesco for helpful discussions, and Monica Beltrame for her encouragement and critical suggestions. We would also like to thank the anonymous referee who substantially improved our algebra. This work was partially supported by a grant from Ministero della Ricerca Scientifica e Tecnologica (MURST) of Italy.

References

Bell,L.R. and Byers,B. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3445–3449.
Bell,S.P., Pikaard,C.S., Reeder,R.H. and Tjian,R. (1989) *Cell*, **59**, 489–497.
Berta,P., Hawkins,J.R., Sinclair,A.H., Taylor,A., Griffiths,B.L., Goodfellow,P.N. and Fellous,M. (1990) *Nature*, **348**, 448–450.
Bhattacharyya,A., Murchie,A.I.H., von Kitzing,E., Diekmann,S., Kemper,B. and Lilley,D.M.J. (1991) *J. Mol. Biol.*, **221**, 1191–1207.
Bianchi,M.E. (1988) *EMBO J.*, **7**, 843–849.
Bianchi,M.E. (1991) *Gene*, **104**, 271–275.
Bianchi,M.E., Beltrame,M. and Paonessa,G. (1989) *Science*, **243**, 1056–1059.
Bianchi,M.E., Beltrame,M. and Falciola,L. (1992a) In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology*, Vol. 6. Springer Verlag, Berlin, pp. 112–128.
Bianchi,M.E., Falciola,L., Ferrari,S. and Lilley,D.M.J. (1992b) *EMBO J.*, **11**, 1055–1063.
Bruhn,S.L., Pil,P.M., Essigman,J.M., Housmann,D. and Lippard,S.J. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 2307–2311.
Bustin,M., Lehn,D.A. and Landsman,D. (1990) *Biochim. Biophys. Acta*, **1049**, 231–243.
Drlica,K. and Rouvière-Yaniv,J. (1987) *Microbiol. Rev.*, **51**, 301–319.
Duckett,D.R., Murchie,A.I.H., Diekmann,S., von Kitzing,E., Kemper,B. and Lilley,D.M.J. (1988) *Cell*, **55**, 79–89.
Fisher,R.P., Parisi,M.A. and Clayton,D.A. (1989) *Genes Dev.*, **3**, 2202–2217.
Gartenberg,M.R. and Crothers,D.M. (1988) *Nature*, **333**, 824–829.
Giese,K., Amsterdam,A. and Grosschedl,R. (1991) *Genes Dev.*, **5**, 2567–2578.
Giese,K., Cox,J. and Grosschedl,R. (1992) *Cell*, **69**, 1–20.
Gubbay,J., Collignon,J., Koopman,P., Capel,B., Economou,A., Münsterberg,A., Vivian,N., Goodfellow,P. and Lovell-Badge,R. (1990) *Nature*, **346**, 245–250.
Harley,V.R., Jackson,D.I., Hextall,P.J., Hawkins,J.R., Berkovitz,G.D., Sockanathan,S., Lovell-Badge,R. and Goodfellow,P.N. (1992) *Science*, **255**, 453–456.
Horikoshi,M., Bertuccioli,C., Takada,R., Wang,J., Yamamoto,T. and Roeder,R.G. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 1060–1064.
Horwitz,M.S.Z. and Loeb,L.A. (1988) *Science*, **241**, 703–705.
Jantzen,H.-M., Admon,A., Bell,S.P. and Tjian,R. (1990) *Nature*, **344**, 830–836.
Kelly,M., Burke,J., Smith,M., Klar,A. and Beach,D. (1988) *EMBO J.*, **7**, 1537–1547.
Kerppola,T.K. and Curran,T. (1991a) *Science*, **254**, 1210–1214.
Kerppola,T.K. and Curran,T. (1991b) *Cell*, **66**, 317–326.
Kim,J., Zwieb,C., Wu,C. and Adhya,S. (1989) *Gene*, **85**, 15–23.
Lee,D.K., Horikoshi,M. and Roeder,R.G. (1991) *Cell*, **67**, 1242–1250.
Levene,S.D. and Zimm,B.H. (1989) *Science*, **245**, 396–399.
Lilley,D.M.J. (1990) In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology*, Vol. 4. Springer Verlag, Berlin, pp. 55–77.
Lilley,D.M.J. (1992) *Nature*, **357**, 282–283.
Liu-Johnson,H.-N., Gartenberg,M.R. and Crothers,D.M. (1986) *Cell*, **47**, 995–1005.
McStay,B., Frazier,M.W. and Reeder,R.H. (1991) *Genes Dev.*, **5**, 1957–1968.
Murchie,A.I.H., Clegg,R.M., von Kitzing,E., Duckett,D.R., Diekmann,S. and Lilley,D.M.J. (1989) *Nature*, **341**, 763–766.

Nasrin,B., Buggs,C., Fu Kong,X., Carnazza,J., Goebl,M. and Alexander-Bridges,M. (1991) *Nature*, **354**, 317–320.
Panayotatos,N. and Fontaine,A. (1987) *J. Biol. Chem.*, **262**, 11364–11368.
Parisi,M.A. and Clayton,D.A. (1991) *Science*, **252**, 965–969.
Pikaard,C.S., Pape,L.K., Henderson,S.L., Ryan,K., Paalman,M., Lopata,M.A., Reeder,R.H. and Sollner-Webb,B. (1990a) *Cell. Mol. Biol.*, **10**, 4816–4825.
Pikaard,C.S., Smith,S.D., Reeder,R.H. and Rothblum,L. (1990b) *Cell. Mol. Biol.*, **10**, 3810–3812.
Pil,P.M. and Lippard,S.J. (1992) *Science*, **256**, 234–236.
Seeman,N.C., Rosenberg,J.M. and Rich,A. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 804–808.
Sinclair,A.H., Berta,P., Palmer,M.S., Hawkins,J.R., Griffiths,B.L., Smith,M.J., Foster,J.W., Frischauf,A.-M., Lovell-Badge,R. and Goodfellow,P.N. (1990) *Nature*, **346**, 240–244.
Singh,J. and Dixon,G.H. (1990) *Biochemistry*, **29**, 6295–6302.
Staben,C. and Yanofsky,C. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4917–4921.
Starr,D.B. and Hawley,D.K. (1991) *Cell*, **67**, 1231–1240.
Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1991) *Methods Enzymol.*, **185**, 60–89.
Sugimoto,A., Iino,Y., Maeda,T., Watanabe,Y. and Yamamoto,J.W. (1991) *Genes Dev.*, **5**, 1990–1999.
Tabor,S. and Richardson,C.C. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1074–1078.
Thompson,J.F. and Landy,A. (1988) *Nucleic Acids Res.*, **16**, 9687–9705.
Travis,A., Amsterdam,A., Belanger,C. and Grosschedl,R. (1991) *Genes Dev.*, **5**, 880–894.
Tremethick,D.J. and Molloy,P.L. (1986) *J. Biol. Chem.*, **261**, 6986–6992.
Tremethick,D.J. and Molloy,P.L. (1988) *Nucleic Acids Res.*, **16**, 1471–1486.
van de Wetering,M. and Clevers,H. (1992) *EMBO J.*, **11**, 3039–3044.
van de Wetering,M., Oosterwegel,M., Dooijes,D. and Clevers,H. (1991) *EMBO J.*, **10**, 123–132.
Waterman,M.L. and Jones,K.A. (1990) *New Biol.*, **2**, 621–636.
Waterman,M.L., Fischer,W.H. and Jones,K.A. (1991) *Genes Dev.*, **5**, 656–669.
Watt,F. and Molloy,P.L. (1988) *Nucleic Acids Res.*, **16**, 1471–1486.
Wright,J.M. and Dixon,G.H. (1988) *Biochemistry*, **27**, 576–581.
Wu,H.-M. and Crothers,D.M. (1984) *Nature*, **308**, 509–513.
Yang,C.-C. and Nash,H.A. (1989) *Cell*, **57**, 869–880.

Received on July 7, 1992; revised on August 26, 1992